基于二分搜索密度峰算法的 RBFNN 在月降水预报中的应用 *

蒋林利 [1,2]，吴建生 [1]，丁立新 [2]

(1. 广西科技师范学院 数学与计算机科学学院，广西 来宾 546199; 2. 武汉大学 软件工程国家重点实验室，武汉 430070)

摘 要：针对径向基函数(RBF）网络结构和初始数据中心难以客观确定的不足，采用二分搜索密度峰值聚类算法（TSDPCA）找到数据中心值及数据簇类个数作为 RBF 神经网络的初始参数和隐藏层节点数，再利用梯度下降法优化 RBFNN 结构及各个参数建立预报模型，并应用于广西月降水预报中，以检验该模型的有效性。结果表明，与 K-RBFNN 和 OLS-RBFNN 的模型相比，TSDPCA-RBFNN 预报平均相对误差值下降了 10%~35%，具有更好的预报性能。

关键词：二分法；密度峰；径向基神经网络；降水预报；梯度下降法

中图分类号：TP183　　doi: 10.3969/j.issn.1001-3695.2017.07.0708

Application of RBFNN based on TSDPCA in monthly runoff forecasting

Jiang Linli[1,2], Wu Jiansheng[1], Ding Lixin[2]

(1. *School of Mathematics & Computer Science Guangxi Science & Technology Normal University*, *Laibin Guangxi* 546199, *China*; 2. *State Key Laboratory of Software Engineering*, *Wuhan University*, *Wuhan* 430072 *China*)

Abstract: Aiming at the shortages of the RBFNN network structure and initial data center being difficult to objectively determined, this paper puts forward using two searching density peak clustering algorithm (TSDPCA) to find data center value and number as the initial parameters of RBFNN and the number of hidden layer nodes quickly, and finally using gradient descent method of optimization the structure of RBFNN and various parameters to set up precipitation forecast model, which was applied in forecasting monthly rainfall of Guagnxi, aiming to testing the effectiveness of the model. The result shows that the mean relative error of TSDPCA-RBFNN forecasting was decreased by 10%-35% compared with K-RBFNN and OLS-RBFNN model, which has better prediction performance.

Key Words: Dichotomy; Density Peak; RBFNN; Rainfall Forecasting; Gradient descent method

0 引言

长期降水预报是对一个月或一个季度的降雨量的预测，在防洪防灾、水库调度和农业灌溉等各方面都有很重要的意义，特别是在农业气象上，准确率高的长期降水预报能够给经济产生巨大的效益。由于广西地形特殊复杂，天气气候多变。在长期天气预报中，主要采用的是线性统计预测方法，而这些方法在处理复杂的非线性降雨问题时具有一定的局限性[1]。因此，深入研究气象长期降水预报的新技术，提高实际预报的精确度，是一项十分重要的研究课题。

近年来，神经网络方法在气象预测中显示出了更好的预报能力[2-5]。这为大气科学领域的气候分析和气候预报提供了新的技术和方法。其中应用最多的为 BP 神经网络和径向基神经网络(radial basis function neural network，RBFNN)。BP 算法是最普通、最典型的神经网络训练算法，但基于梯度下降法的 BP 算法要依赖初始权值的选择，再加上实际问题则是极其复杂的多维曲面，并存在多个局部极值点，所以 BP 网络算法存在收敛速度慢和易陷入局部最优。RBFNN 是一种结构简单、运算速度快和很强的非线性映射功能的三层前馈神经网络，更加适合用于具有非线性时间序列的预测问题[6~10]。但 RBFNN 的性能主要受到其本身拓扑结构因素的影响，即隐藏层基函数的中心、隐藏层的个数和扩展宽度这三个参数难以确定。到目前为止，还没找到一种合适的方法来确定以上参数。很多研究者将聚类方法的思想应用优化 RBFNN 结构设计中，提出了基于 K 近邻、K-均值聚类、K-中心聚类、正交最小二乘法（OLS）和自组织映射（SOM）聚类等拓扑结构的设计方法优化 RBFNN 网络结构，并取得了一系列的成果[11~15]。2014 年 6 月，Rodriguez 等人[16]提出了一种新的聚类方法，基于相互距离，即利用欧氏距

基金项目：广西自然科学基金资助项目（2014GXNSFAA118027）；广西高校科学技术研究项目（YB2014467）；广西高校中青年教师基础能力提升项目（2017KY0896）；广西重点学科资助项目（070105）

作者简介：蒋林利（1976-），女，广西全州人，副教授，硕士，武大学访问学者，主要研究方向为机器学习、数据挖掘等；吴建生（1974-），男，陕西咸阳人，教授，硕士，主要研究方向为人工智能神经网络；丁立新（1967-），男，博导，主要研究方向为智能信息处理、机器学习等研究.

离计算样本数据点间的距离，通过局部密度和到更高密度点的最短距离这两个属性来构造聚类的决策图，然后操作者选取这两个属性值都大的数据点作为聚类中心。该方法具有简洁、计算速度快、能识别各种形状的簇类等优点。但针对多维复杂的数据时，采用欧式距离度量相似性存在鲁棒性和物理概念不明确等不足以及存在聚类不稳定性。文献[17]提出采用密度峰改进的 K-medoids 聚类获得新的聚类准则函数，提高了聚类的精度和收敛速度，同时对噪声有较好的鲁棒性。文献[18]为了解决 K-中心等聚类方法易陷入局部最优解的问题采用密度信息对其进行改进并应用雷暴分类中，具有精确度高和收敛速度快等优点。文献[19]利用密度参数改进 K-均值算法对样本数据进行聚类获取中心作为 RBF 基函数的中心，其预报精度均高于标准的 K-均值 RBFNN 预测模型和 BP 网络预测模型，但 RBF 网络的节点个数 K 值是根据经验公式和网络试算进行长时间和大量的实验摸索后确定的，能找到较合适的网络模型，而且在很大程度上依赖于具体处理的问题，同时也需要使用者具有神经网络和问题领域两方面的先验知识。还没有找到一种能自动地获取网络节点个数和中心点的方法。

本文针对上述的不足，提出一种二分搜索密度峰聚类算法（two search density peak clustering algorithm,TSDPCA）自动快速地获取聚类中心和聚类个数来优化 RBF 神经网络结构。首先，利用余弦夹角相似度计算样本数据间距离，更多地注重了维度间相对层面的差异，克服了欧氏距离存在的缺陷，同时具有平移和旋转不变性的优点；其次，根据用户设定的 d_c 值计算数据集的局部密度 ρ_i 和 δ_i (其对应的距离)，$\lambda_i=\rho_i.\delta_i$ 对进行排序取最大的两个点作为聚类中心点，按离最近邻的高密度的最小距离规划到相应的两个类簇中，将第二个类簇重复上述方法继续分为两个类簇，直到相邻两个簇之间变化很小时，停止继续分类，并保存聚类的中心值和聚类的个数；最后利用 TSDPCA 方法和梯度下降法化混合优化 RBFNN 参数和结构建立预报模型，并应用于广西桂林地区月降水量的预测，结果表明，网络收敛速度快、预报精确度较高。

1　径向基函数神经网络

1.1　RBFNN 结构

RBFNN 是一种由输入层、隐藏层和输出层三层构成的前馈神经网络[20,21]。其网络拓扑结构如图 1 所示。其基本原理是用 RBF 作为隐藏层单元的“基”，隐藏层对输入数据进行转换，将低维的输入数据自动地转换到高维的空间内。RBFNN 的激活函数采用高斯函数，可用如下数学公式表示为

$$\varphi(X_N)=\exp(-\frac{\| X_N-C_j \|}{2\sigma^2}),(N=1,2,\cdots N;j=1,2,\cdots k) \quad (1)$$

其中：$X_N=(X_1,X_2\cdots,X_n)$ 是输入数据；$\left\|X_N-C_j\right\|^2$ 是欧氏范数；C_j 是高斯函数中心；σ 是高斯函数方差。

对于输入数据样本 X_N，则 RBFNN 的期望输出结果可用数学方式表示为

$$Y_i'=F(X_i)=\sum_{j=1}^{k}\omega_{ij}\exp(\frac{-\left\|X-C_j\right\|^2}{2\sigma^2}) \quad (2)$$

其中：ω_{ij} 是隐藏层到输出层的连接权值。

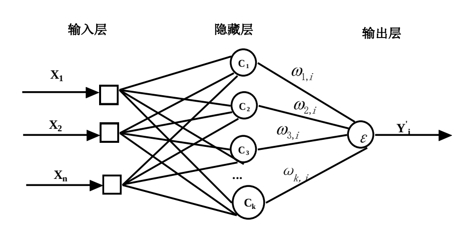


图 1　RBF 神经网络结构

1.2　RBFNN 梯度下降学习法

通过最小化目标函数采用梯度下降法实现对各隐节点数据中心、隐藏层的扩展常数以及隐层到输出层间权值的调整。该网络学习的目标函数为

$$E=\frac{1}{2}\sum_{i=1}^{N}\beta_i e_i^2 \quad (3)$$

$$e_i=Y_i-Y_i'=Y_i-F(X_i) \quad (4)$$

其中：β_i 为遗忘因子，Y_i 为 X_i 对应的真实值，e_i 为真实值与期望输出值的误差。

中心 C_i、宽度 γ_i、权值 ω_i 学习迭代方法如下：

$$C_i=C_i+\Delta_{C_i}F(X)=C_i+\frac{2\omega_i}{\gamma^2}\varphi_i(X)\left\|X-C_i\right\| \quad (5)$$

$$\gamma_i=\gamma_i+\Delta_{\gamma_i}F(X)=\gamma_i+\frac{2\omega_i}{\gamma^3}\varphi_i(X)\left\|X-C_i\right\|^2 \quad (6)$$

$$\omega_i=\omega_i+\Delta_{\omega_i}F(X)=\omega_i+\varphi_i(X) \quad (7)$$

2　基于二分搜索密度峰的聚类算法的 RBFNN 模型

2.1　二分搜索密度峰的聚类算法

文献[19]提出基于密度峰值和距离的快速聚类算法，使用的距离样本数据集是所有点及样本数据两点间距离构成的三维矩阵。其利用的欧式距离计算数据点间的距离，主要体现数值的绝对差异，不能很好的体现维度之间的差异。本文利用夹角余弦相似度计算数据集点间的距离，对绝对的数值不敏感，而更多的从方向上或维度间相对层面区分差异，采用 n 个 d 维的样本数据集 $X_{nd}=(X_{1d},X_{2d}\cdots X_{nd})$ 利用夹角余弦距离构造如下三维矩阵的距离样本集 $S_{n*(n-1)/2}$。夹角余弦数学表示如下：

$$D(X_i,X_j)=\frac{\sum_{a=1}^{d}(X_{ia}X_{ja})}{\sqrt{\sum_{a=1}^{d}X_{ia}^2\sum_{a=1}^{d}X_{jd}^2}} \quad (8)$$

$$S_{n*(n-1)/2}=\begin{bmatrix} 1 & 2 & D(X_1X_2) \\ \cdots & \cdots & \cdots \\ 1 & n & D(X_1X_n) \\ 2 & 3 & D(X_2X_3) \\ \cdots & \cdots & \cdots \\ n-1 & n & D(X_{n-1}X_n) \end{bmatrix} \quad (9)$$

其聚类算法有两个前提假设，一是聚类中心的局部密度高于其周围的样本点的局部密度；二是聚类中心与比其局部密度还高的点间的距离相对较大。其聚类效果可以检测出任意形状的类簇。该聚类方法根据密度和距离构画出的决策分布图，将选择明显远离数据样本集大部分的数据样本点且位于决策分布图右上角区域的密度峰作为初始聚类中心；按离最近邻的高密度的最小距离规划到相应的类簇中。其算法虽然简单有效，但却需要人工选取聚类中心，这样对于同一批数据可能会因人不同选取的中心也不同，同时可能会出现一个类被错误地分成多个簇类的特殊情况。 为了避免多余的簇类中心，本文综合考虑 ρ_i 和 δ_i 的值。提出相邻聚类间的变化值小于 0.01 时停止搜索密度峰聚类。本文用到的参数和实现聚类的步骤如下：

局部密度定义为[22]

$$\rho_i = \sum_j X(d_{ij} - d_c) \tag{10}$$

其中: $X(x)=\begin{cases}1 & (x<0)\\ 0 & (x\ge 0)\end{cases}$，这个公式说明找到与第 i 个样本数据点间的距离小于 d_c 的样本数据点的个数；其中 d_c 为截断距离，对于选取一个 d_c 的原则是使得每个数据点的“邻居“个数均约为数据点总数的 1%-3%。本文设定的值为式(9)所有点的相互距离中由小到大排序占总数 2%位置的距离数值。

与高密度点间的距离定义为

$$\delta_i = \min_{j:\rho_j>\rho_i}(d_{ij}) \tag{11}$$

即找到比第 i 个样本数据点的局部密度都大的所有样本数据点中，与第 i 个样本数据点间的距离的最小值。如果对于具有最大密度的样本数据点，通常取 $\delta_i = \max_{j>2}(d_{ij})$ 。为了统一标准，将 δ_i 的值进行归一化[0,1]。

$$\lambda_i = \rho_i . \delta_i \tag{12}$$

聚类相邻之间差值为

$$\ell_{ij} = \lambda_i - \lambda_j \le 0.01 \tag{13}$$

其中： λ_i 值越大，其对应的数据样本点越可能是数据中心点。因此对 λ_i 进行排序，取 λ_i 最大两个值的数据样本点作为聚类中心，然后按离最近邻的高密度的最小距离规划到相应的两个类簇中，聚类后保存第一类簇的聚类中心及簇类数据样本,将另一个聚类数据样本继续按上述的方法聚类，直到相邻间的变化很小即满足 ℓ_{ij} 条件为止，最终获得的类簇数目个数和类簇中心位置值。

具体步骤如下：

输入：实验样本数据集 $X_{nd}=(X_{1d},X_{2d}\cdots X_{nd})$ 。

输出：聚类结果中心值 $C=(C_{1d},C_{2d}\cdots C_{kd})$,k 为聚类个数。

a)按照式（8）计算任意两数据样本之间的距离 d_{ij} ，构造出式（9）的三维矩阵数据集 $S_{n*(n-1)/2}$ ；

b)分别按照式（10）（11），计算每个数据样本的局部密度 ρ_i 及该点到具有更高局部密度的数据样本的最小距离 δ_i ；

c)根据式（12）计算 λ_i 值，并进行降序排列，取 λ_i 最大两个值的数据样本作为两个聚类中心，并按离最近邻的高密度的最小距离将样本数据规划到相应的类簇中，并保存第一类聚类中心及对应的位置；

d)计算相邻之间的误差 ℓ_{ij} ；

e)将另一个聚类的样本数据集重复执行步骤 b）~d）,直到满足式（13）的结束条件为止；

f)获得聚类中心 C_k 及聚类个数，将聚类中心的位置值作为下标取出其对应的 X_{nd} 样本中的数据样本。

将上述方法得到的聚类中心 C_k 和聚类数目作为RBFNN 网络的基函数中心和隐节点个数。

2.2　基于 TSDPCA 的 RBFNN 建立预测模型

RBFNN 模型的三个参数，输入神经元个数采用 10 个降水因子作为 RBFNN 的输入神经元节点个数；隐藏神经元个数因到目前为止还没有找到准确、科学的、有效的确定方法，大多是通过数值实验对比和经验公式来确定，本文采用上述聚类获取的聚类中心个数作为该模型的隐藏神经元的个数；隐节点的扩展常数或宽度值，可通过计算隐节点数据中心之间的最小距离，数学公式表示为

$$SP = \min(d_{ij}) = \sqrt{\sum_{k=1}^{K}(C_{ik}-C_{jk})^2} \tag{14}$$

建立 TSDPCA 的 RBFNN 模型简单可归纳为四个步骤：

a)通过均生函数和奇异谱处理的降水因子，再利用主成分分析提取综合降水预报因子并进行归一化处理，提取出主要预报因子作为 RBFNN 训练输入数据样本矩阵；

b)利用夹角余弦计算数据样本点之间的距离并构造 3 维矩阵；

c)采用本文描述的 TSDPCA 聚类方法找出数据中心值和中心数目；

d)通过式（14）计算出隐层的扩展宽度，然后采用梯度下降的高斯函数的 RBF 建立预报模型。

基于 TSDPCA 的 RBFNN 模型建立过程如图 2 所示。

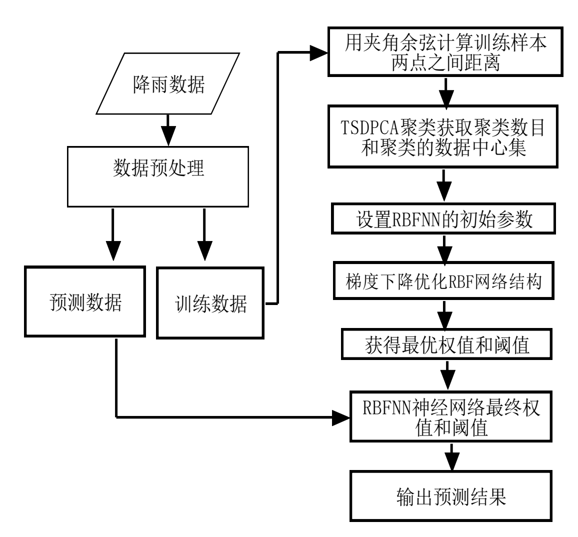


图 2　基于 TSDPCA 的 RBFNN 建立预测模型

3　实验与分析

3.1　研究区域和数据

现以广西桂林气象局提供的十个站点每月的实时降雨数据（1949—2008 年）为例，建立基于 TSDPCA-RBNN 的预测模型。其中，训练数据总长度为 720 个，影响降雨因子 360 个，其中选取 1949—2005 年的 684 个月数据作为训练集，2006-2008 年 36 个月数据作为测试集。

3.2　数据预处理

本文对 360 个降水因子是通过采用奇异谱分析法(singular spectrum analysis, SSA)对降水原始序列重构，并利用均生函数（mean generating function, MGF）对其进行延拓[20]，再采用 80%的主成分分析法提取最强的降水综合因子 10 个，作为 RBFNN 的输入矩阵向量，因 RBFNN 的输入、输出数据要求在[0,1]，最后将上述处理后的降水样本进行归一化处理。

3.3　参数设置

本程序运行的环境是在 Win7 系统，MATLAB 7 软件，实验中 d_c 的初始“邻居“个数值设为 2%(在 1%-3%之间取值对结果影响不大)，本程序 TSDPCA 自动获取的数据中心个数为 14 个,RBFNN 训练的隐节点数据中心学习系数 lrCent=0.01;隐节点扩展常数学习系数 lrSP = 0.01;隐节点输出权值学习系数 lrW=0.01;隐层节点个数为聚类的数据中心个数;目标误差=0.9;最大训练次数=5000。

3.4　结果分析

图 3 为广西桂林站 1949 年到 2008 年每个月的平均降水量，本文从中可以发现一年中平均降雨量最大的值为 250mm 左右，最少降水雨量值为 50mm 左右。通过实际对比如果月降水量大于或等于 250mm 将会出现特大洪灾，如果小于或等于 50mm 将会出现干旱灾害。

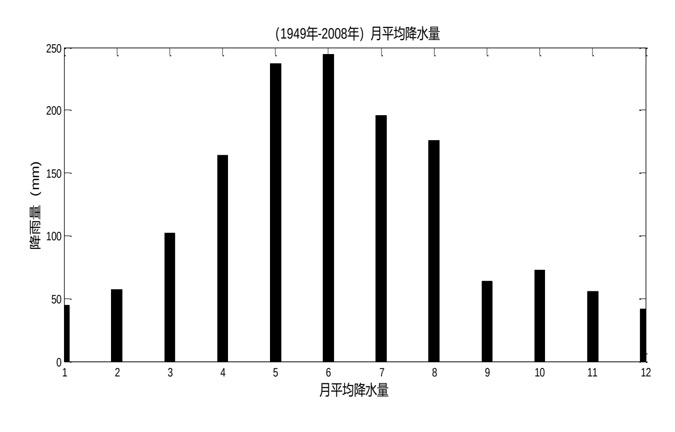


图 3　月平均降水量（1949 年-2008 年）

为了更好地考察效果，本文的 TSDPCA-RBFNN 模型与 K-均值 RBFNN 模型、正交最小二乘 RBFNN 模型（OLS-RBFNN）对月降水预报的结果进行对比（说明：这里的三种方法中提到的 RBFNN 都是通过梯度下降法优化网络结构）。同时这里还引入表 1 所示的五种评价指标。

表 1　五种性能评价指标及公式

指标	公式
平均绝对误差（MAE）	$MAE = \frac{1}{N}\sum_{i=1}^{N}\left\|\left\|Y_i - Y_i'\right\|\right\|$， Y_i is the actual value , Y_i' is the predicted value
平均相对误差 (MAPE)	$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left\|Y_i - Y_i'\right\|}{\left\|Y_i\right\|}*100$
均方根误差 (RMSE))	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|Y_i - Y_i'\right\|^2}$
洪灾：（error>250mm）	$F = \sum_{i=1}^{N} I_i$， $I_i = \begin{cases} 1 & \|Y_i - Y_i'\| \geq 250 \\ 0 & \|Y_i - Y_i'\| < 250 \end{cases}$
旱灾：(error<50mm)	$D = \sum_{i=1}^{N} L_i$， $L_i = \begin{cases} 1 & \|Y_i - Y_i'\| < 50 \\ 0 & \|Y_i - Y_i'\| \geq 50 \end{cases}$

图 4~6 是分别是三种不同预报模型对 2006—2008 年的月降水预报结果与实际观察数据进行对比，从实验中可以发现三种模型的模拟趋势基本跟实际观察值的趋势基本吻合，从图 4~6 预测图可以看出各种模型对降水的预报能力， TSDPCA-RBFNN 的预报结果是最好的，它跟实际观察值的趋势基本一致，偏离是最小的，具有良好的一致性。

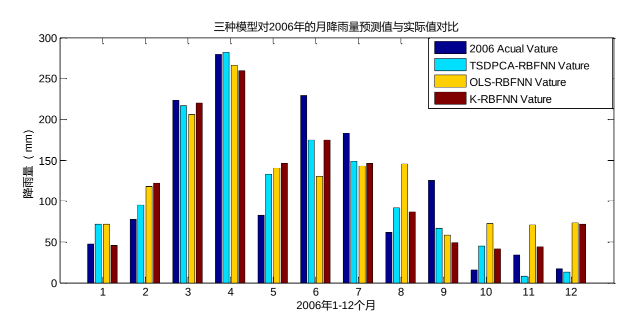


图 4　三种模型对 2006 年的降雨量预测值

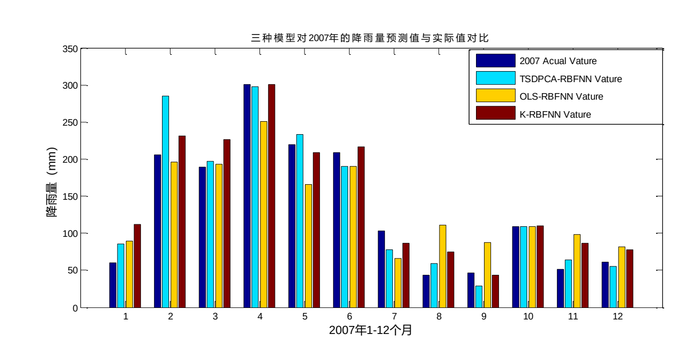


图 5　三种模型对 2007 年的降雨量预测

图 7 是利用 TSDPCA 根据 λ_i 降序对数据集值进行分类，从图中发现聚类数目达到 14 时 λ_i 变化几乎趋于稳定，说明分为 14 个类簇是比较合适的。根据相邻之间变化比较小作为终止条件可以自动地获取聚类数目和中心点。

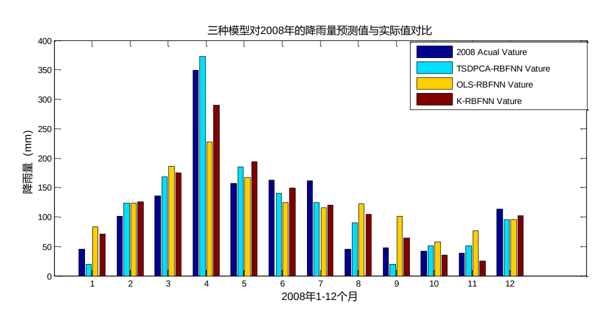


图 6　三种模型对 2008 年的降雨量预测值

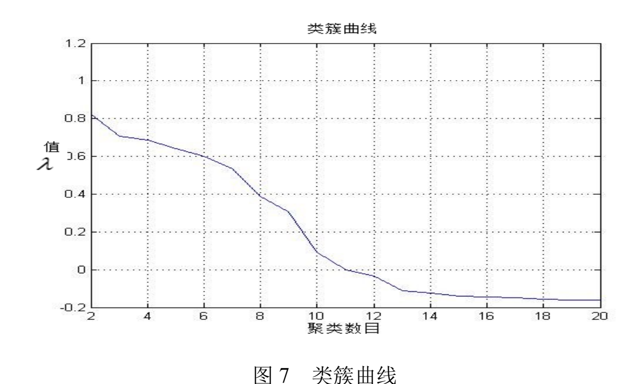


图 7　类簇曲线

图 8、9 利用训练数据在训练过程中误差达到平稳的收敛次数，从图中可以看出 TSDPCA-RBFNN 模型的收敛次数值要比其他两种模型收敛次数值小。

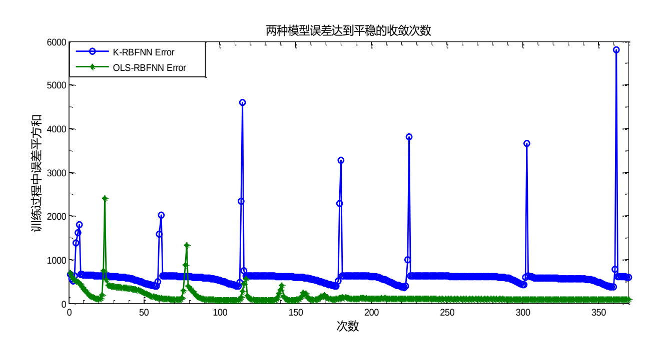


图 8　两种模型误差达到平稳的收敛次数

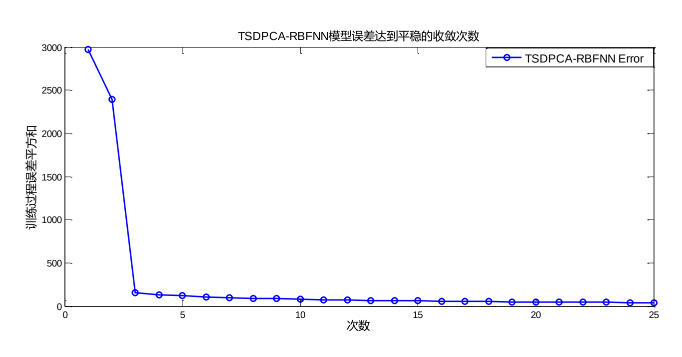


图 9　TSDPCA-RBFNN 模型误差达到平稳的收敛次数

表 2 给出了三种模型在训练过程中拟合的时间以及对 2006 年到 2008 年这三年的月平均降雨量在三种不同模型中预测误差的精确度，可以看出，TSDPCA-RBFNN 平均相对误差、均方根误差和平均绝对误差都是最小的，比 K-RBFNN 模型的平均相对误差下降了 10%左右，比 OLS-RBFNN 模型下降了 35%。这就说明 TSDPCA-RBFNN 的预测精度有很大的提高，能较准确的预测降水预报。此外，本文对比了三种模型的收敛速度，主要从模型训练拟合时间和误差达到平稳的训练收敛次数两个方面考虑，TSDPCA-RBFNN 模型的训练拟合时间为 18.6345 秒，收敛次数 10 次；而 OLS-RFBNN 模型的训练拟合时间为 61.6801 秒，收敛次数 180 次；K-RBFNN 模型的训练拟合时间为 44.6128 秒，收敛次数 362 次。由此可见，TSDPCA-RBFNN 模型比其他两种模型的收敛速度快。

表 2　三种降水预测模型的性能统计

模型	K-RBFNN	OLS-RBFNN	**TSDPCA-RBFNN**
		训练数据	
训练拟合时间（s）	44.6128	61.6801	**18.6345**
		测试数据（2006--2008）	
平均相对误差（MAPE）	42.238%	67.076%	**32.84%**
均方根误差（RMSE）	44.3435	49.2310	**29.3636**
平均绝对误差（MAE）	27.75071	41.59739	**23.9672**

如果以超过 250 mm 的月降雨量为特大洪灾来预测，从图 4 中可以发现，2006 年实况发生特大洪灾是 4 月，发生旱灾是 1 月、10 月、11 月和 12 月，TSDPCA-RBFNN 模型中预测的特大洪灾显示为 4 月及旱灾为 1 月、10 月、11 月和 12 月，与实际情况完全吻合。从图 5 发现 TSDPCA-RBFNN 模型预测 2007 年特大洪灾为 2 月、4 月和 5 月，旱灾为 8 月和 9 月，2007 年实况特大洪灾是 4 月，旱灾是 8 月和 9 月，再仔细对比发现，2007 年的 2 月和 5 月实况降雨量超过 200MM 以上，说明雨水量较多，所以该模型预测效果几乎跟实况基本吻合。从图 6 发现 2008 年的预测特大洪灾与实况都是 4 月，预测的旱灾与实况旱灾也比较接近。从这三个图中可以发现，相比其他两个预测模型，本文提出的 TSDPCA-RBFNN 预测模型的精确度更高，预报更稳定。该模型能为防洪防旱灾害和径流预报等提供一个重要的参考价值，可作为一种较好的预报工具。

4　结束语

降水因受到众多不同因素的综合影响，而且具有非线性的典型特征，而 RBFNN 具有较强的非线性关系，以及参数少、自学习和自适应等能力,能适用于这类气象因素的建模和预报。但在实际的应用研究中，一般都是采用手工试凑的方法来进行设计 RBF 算法结构和调整训练参数。需要大量耗时的实验探索才能找到合适的网络结构。针对 RBFNN 网络结构和初始化参数难以客观确定的不足,特别是隐藏层的基函数中心的选取问题对整个 RBF 网络的是否成功的关键所在。本文提出的 TSDPCA 方法能自动快速、准确的获取聚类中心和聚类个数，克服了 RBFNN 不足。结合梯度下降法优化 RBFNN 结构建立预测模型。通过实验研究，得出以下结论：

a）从 RBFNN 网络的收敛速度来看，与传统的 K 均值 RBFNN 模型和 OLS-RBFNN 模型比较，基于 TSDPCA 的 RBFNN 模型对确定簇类数目和基函数中心等方面有一定的优势，同时网络的收敛速度和预测精度都有比较明显的提高。

b）对于广西月降水预报而言，改进的 RBFNN 模型预报在月降水量的平均相对误差的精度上有较大的提高，2006，2007，2008 年三年对洪灾旱灾的预测几乎跟实际完全吻合，为在不确定因素非线性较强的长期降水预报提供了一种新的参考方法。

c)本文应用的 TSDPCA 优化 RBFNN 对月降水量建立预测模型时，采用余弦相似度公式代替欧式距离公式来计算降水数据样本间的距离，可进一步分析研究数据之间的相似性，以及综合一些智能优化算法优化该模型网络的隐节点数，以更进一步提高 RBFNN 模型的预测精度；还有就是 d_c 阈值没有确定的方法，有待进一步研究。

参考文献：

[1] 金龙，罗莹，林振山. Comparison of long-term forecasting of june-august rainfall over changjiang-huaihe valley [J]. 大气科学进展, 1997, 14 (1): 88-93.

[2] Jiang Linli, Wu Jiansheng. Hybird evolutionary algorithms for artificial neural network training in rainfall forecasting [C]// Proc of International Conference on Advances in Neural Networks. 2013: 191-195.

[3] Wu J, Long J, Liu M. Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm [J]. Neurocomputing, 2015, 148: 136-142.

[4] 胡健伟，周玉良，金菊良. BP 神经网络洪水预报模型在洪水预报系统中的应用 [J]. 水文, 2015, 35 (1): 20-25.

[5] 闵晶晶，孙景荣，刘还珠，等. 一种改进的 BP 算法及在降水预报中的应用 [J]. 应用气象学报, 2010, 21 (1): 55-62.

[6] Yaseen Z M, El-Shafie A, Afan H A, et al. RBFNN versus FFNN for daily river flow forecasting at Johor River, Malaysia [J]. Neural Computing and Applications, 2016, 27 (6): 1-10.

[7] Cao C, Wang S, Tang H. Rainfall forecasting using RBFNN optimized by FKCN [J]. Journal of Hefei University of Technology, 2000.

[8] He X, Guan H, Qin J. A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall [J]. Journal of Hydrology, 2015, 527 (17): 88-100.

[9] 蒋林利. 基于混合优化的 RBF 神经网络集成的降水预报模型 [J]. 柳州师专学报, 2012, 27 (2): 113-119.

[10] 龙文，梁昔明，龙祖强，等. 基于混合进化算法的 RBF 神经网络时间序列预测 [J]. 控制与决策, 2012, 27 (8): 1265-1268.

[11] 谢娟英，高瑞. Num-近邻方差优化的 K-medoids 聚类算法 [J]. 计算机应用研究, 2015, 32 (1): 30-34.

[12] 沈艳，杨春雷，张庆国，等. 基于 RBF 神经网络的池州市降水序列预测 [J]. 安徽农业大学学报, 2012, 39 (3): 451-455.

[13] 刘博，肖长来，梁秀娟. SOM RBF 神经网络模型在地下水位预测中的应用 [J]. 吉林大学学报：地球科学版, 2015, 45 (1): 225-231.

[14] 伊燕平，卢文喜，张耘，等. 基于径向基函数神经网络的地下水数值模拟模型的替代模型研究 [J]. 水土保持研究, 2012, 19 (4): 265-269.

[15] 周炜奔，石跃祥. 基于密度的 K-means 聚类中心选取的优化算法 [J]. 计算机应用研究, 2012, 29 (5): 1726-1728.

[16] Rodriguez A, Laio A. Clustering by fast search and find of density peaks [J]. Science, 2014, 344 (6191): 1492-1496.

[17] 谢娟英，屈亚楠. 密度峰值优化初始中心的 K-medoids 聚类算法 [J]. 计算机科学与探索, 2016 (2): 1-16.

[18] 朱晔，冯万兴，郭钧天，等. 一种改进的 k-中心点聚类算法及在雷暴聚类中的应用 [J]. 武汉大学学报：理学版, 2015, 61 (5): 497-502.

[19] 郭皓，邢贞相，付强等. 基于密度参数 K-均值算法的 RBF 网络及其在降水量预测中的应用 [J]. 水土保持研究, 2014, 21 (6): 299-303.

[20] Wu J. Prediction of rainfall time series using modular RBF neural network model coupled with SSA and PLS [C]// Proc of Asian Conference on Intelligent Information and Database Systems. 2012: 509-518.